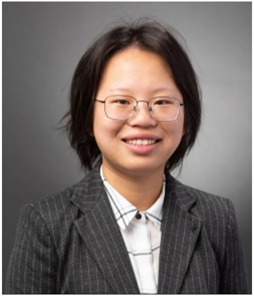
Probabilistic Inequalities
for Sums of Heavy-Tailed Random Matrices

Stas Minsker (University of Southern California)

September 19, 2025

"Mathematical Statistics in the Information Age" Conference

Research supported by DMS CAREER-2045068

based on a joint work in with Moritz Jirak, Yiqiu Shen and Martin Wahl

Concentration of measure

- Concentration of measure phenomenon formalizes the idea that

  $\boxed{\text{nice functions of many independent random variables are “essentially constant”}}$

Concentration of measure

- Concentration of measure phenomenon formalizes the idea that

  $\boxed{\text{nice functions of many independent random variables are “essentially constant”}}$

- This idea can serve as a "bridge" between random and deterministic quantities.

Concentration of measure

- Concentration of measure phenomenon formalizes the idea that

  nice functions of many independent random variables are "essentially constant"

- This idea can serve as a "bridge" between random and deterministic quantities.
- Examples include the Gaussian (Borell-TIS) inequality, bounded difference (McDiarmid's) inequality, Talagrand's inequality for product measures, matrix Bernstein's inequality, etc.

Concentration of measure

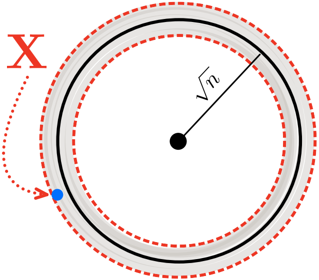
- Concentration of measure phenomenon formalizes the idea that

nice functions of many independent random variables are “essentially constant”

- This idea can serve as a "bridge" between random and deterministic quantities.
- Examples include the Gaussian (Borell-TIS) inequality, bounded difference (McDiarmid’s) inequality, Talagrand’s inequality for product measures, matrix Bernstein’s inequality, etc.
- For example, if $\mathbf{X} = (X_1, \ldots, X_n) \sim N(0, I_n)$ then $\mathbb{E}\|\mathbf{X}\|_2 \in \left[\frac{n}{\sqrt{n+1}}, \sqrt{n}\right]$ and

$$\Big| \|\mathbf{X}\|_2 - \mathbb{E}\|\mathbf{X}\|_2 \Big| \leq \sqrt{2t}$$

with probability at least $1 - e^{-t}$.



Concentration of measure

- For example, if $\mathbf{X} = (X_1, \ldots, X_n) \sim N(0, I_n)$ then $\mathbb{E}\|\mathbf{X}\|_2 \in \left[\frac{n}{\sqrt{n+1}}, \sqrt{n}\right]$ and

$$\left| \|\mathbf{X}\|_2 - \mathbb{E}\|\mathbf{X}\|_2 \right| \leq \sqrt{2t}$$

  with probability at least $1 - e^{-t}$.
- If $\mathbf{W} \in \mathbb{R}^{n \times p}$, $n \geq p$, has i.i.d. normal coordinates, then

$$\left\| \frac{W^T W}{n} - I_p \right\| \leq \sqrt{\frac{p}{n}} + \sqrt{\frac{2t}{n}}$$

  with probability at least $1 - e^{-t}$ ("quantitative version" of Bai-Yin theorem).

Concentration of measure

- For example, if $\mathbf{X} = (X_1, \ldots, X_n) \sim N(0, I_n)$ then $\mathbb{E}\|\mathbf{X}\|_2 \in \left[\frac{n}{\sqrt{n+1}}, \sqrt{n}\right]$ and

$$\Big| \|\mathbf{X}\|_2 - \mathbb{E}\|\mathbf{X}\|_2 \Big| \leq \sqrt{2t}$$

  with probability at least $1 - e^{-t}$.
- If $\mathbf{W} \in \mathbb{R}^{n \times p}$, $n \geq p$, has i.i.d. normal coordinates, then

$$\left\| \frac{W^T W}{n} - I_p \right\| \leq \sqrt{\frac{p}{n}} + \sqrt{\frac{2t}{n}}$$

  with probability at least $1 - e^{-t}$ ("quantitative version" of Bai-Yin theorem).
- Often, a.s. boundedness or exponential integrability assumptions are imposed.

$\boxed{\text{What if the random variables of interest have heavy tails?}}$

Concentration of measure

- For example, if $\mathbf{X} = (X_1, \ldots, X_n) \sim N(0, I_n)$ then $\mathbb{E}\|\mathbf{X}\|_2 \in \left[\frac{n}{\sqrt{n+1}}, \sqrt{n}\right]$ and

$$\Big| \|\mathbf{X}\|_2 - \mathbb{E}\|\mathbf{X}\|_2 \Big| \le \sqrt{2t}$$

  with probability at least $1 - e^{-t}$.
- If $\mathbf{W} \in \mathbb{R}^{n\times p}$, $n \ge p$, has i.i.d. normal coordinates, then

$$\left\| \frac{W^T W}{n} - I_p \right\| \le \sqrt{\frac{p}{n}} + \sqrt{\frac{2t}{n}}$$

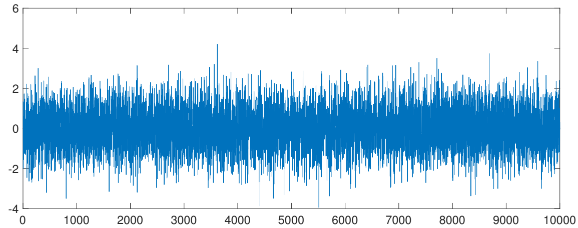
  with probability at least $1 - e^{-t}$ ("quantitative version" of Bai-Yin theorem).
- Often, a.s. boundedness or exponential integrability assumptions are imposed.

What if the random variables of interest have heavy tails?

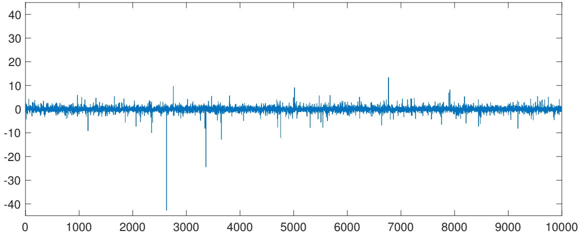
- For the purpose of this talk, a random variable Z has heavy-tailed distribution if

$$\mathbb{E}|Z|^k = \infty$$

  for some $k > 2$.



Standard normal distribution



Student's t-distribution with 3 d.f.

Fuk-Nagaev inequalities

- (Sub)-Gaussian concentration: let $X_1, \ldots, X_n$ be independent, $X_j \sim N(0, \sigma_j^2)$. Then

$$\mathbb{P}\left(\Big| \sum_{j=1}^{n} X_j \Big| \geq t \right) \leq 2 \exp\left(\frac{t^2}{2 \sum_{j=1}^{n} \sigma_j^2} \right)$$

Fuk-Nagaev inequalities

- (Sub)-Gaussian concentration: let $X_1, \ldots, X_n$ be independent, $X_j \sim N(0, \sigma_j^2)$. Then

$$\mathbb{P}\left(\left| \sum_{j=1}^{n} X_j \right| \geq t \right) \leq 2 \exp \left(\frac{t^2}{2 \sum_{j=1}^{n} \sigma_j^2} \right)$$

- Dao Ha Fuk and Sergei Nagaev (1971): let $X_1, \ldots, X_n$ be i.i.d.centered random variables with $p \geq 2$ finite moments. Then

$$\mathbb{P}\left(\left| \sum_{j=1}^{n} X_j \right| \geq t \right) \leq 2 \exp \left(-C_1(p) \frac{t^2}{\sum_{j=1}^{n} \mathbb{E} X_j^2} \right) + \mathbb{P}\left(\max_j |X_j| > t/4 \right)$$
$$+ C_2(p) \left(\frac{\sum_{j=1}^{n} \mathbb{E}|X_j|^p}{t^p} \right)^{2}$$

Fuk-Nagaev inequalities

- (Sub)-Gaussian concentration: let $X_1, \ldots, X_n$ be independent, $X_j \sim N(0, \sigma_j^2)$. Then

$$\mathbb{P}\left(\left| \sum_{j=1}^{n} X_j \right| \geq t \right) \leq 2 \exp\left(\frac{t^2}{2 \sum_{j=1}^{n} \sigma_j^2} \right)$$

- Dao Ha Fuk and Sergei Nagaev (1971): let $X_1, \ldots, X_n$ be i.i.d.centered random variables with $p \geq 2$ finite moments. Then

$$\mathbb{P}\left(\left| \sum_{j=1}^{n} X_j \right| \geq t \right) \leq 2 \exp\left(-C_1(p) \frac{t^2}{\sum_{j=1}^{n} \mathbb{E} X_j^2} \right) + \mathbb{P}\left(\max_j |X_j| > t/4 \right) + C_2(p) \left(\frac{\sum_{j=1}^{n} \mathbb{E}|X_j|^p}{t^p} \right)^2$$

- E. Rio (2017), M. Bakhshizadeh, A. Maleki, V. H. de la Pena (2022) proved bounds implying that

$$\mathbb{P}\left(\left| \sum_{j=1}^{n} X_j \right| \geq t \right) \leq \exp\left(-\frac{t^2}{(2+\delta) \sum_{j=1}^{n} \mathbb{E} X_j^2} \right) + C_2(\delta) p^p \frac{\mathbb{E} \max_j |X_j|^p}{t^p}$$

Fuk-Nagaev inequalities

- Uwe Einmahl and Deli Li (2007); Radek Adamczak (2008) proved that for Banach space-valued centered random variables,

$$\mathbb{P}\left(\Big\|\sum_{j=1}^{n} X_j\Big\| \geq (1+\eta)\mathbb{E}\Big\|\sum_{j=1}^{n} X_j\Big\| + t\right) \leq \exp\left(-\frac{t^2}{(2+\delta)B_n^2}\right) + C(\eta,\delta,p)\frac{\sum_{j=1}^{n}\mathbb{E}\|X_j\|^p}{t^p}$$

where $B_n^2 = \sup_{\|f\|_*=1}\sum_{j=1}^{n}\mathbb{E}\left\langle f, X_j\right\rangle^2$.

Fuk-Nagaev inequalities

- Rio's and Einmahl and Li's inequalities can be used to prove versions of the bounded Law of the Iterated Logarithm.

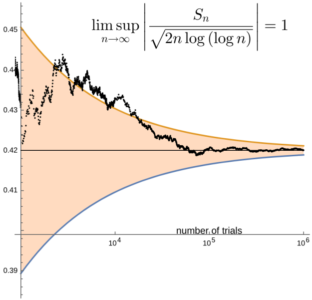


Figure: demonstrations.wolfram.com

Moment inequalities

- Haskell Rosenthal's inequality (1970): let $X_1, \ldots, X_n$ be i.i.d.centered random variables with $p \geq 2$ finite moments. Then

$$\mathbb{E}^{1/p}\Big|\sum_{j=1}^{n} X_j\Big|^p \leq C(p)\left(\left(\sum_{j=1}^{n} \mathbb{E}X_j^2\right)^{1/2} + \left(\sum_{j=1}^{n} \mathbb{E}|X_j|^p\right)^{1/p}\right)$$

Moment inequalities

- Haskell Rosenthal's inequality (1970): let $X_1, \dots, X_n$ be i.i.d.centered random variables with $p \geq 2$ finite moments. Then

$$\mathbb{E}^{1/p}\Big|\sum_{j=1}^{n} X_j\Big|^p \leq C(p)\left(\left(\sum_{j=1}^{n}\mathbb{E}X_j^2\right)^{1/2} + \left(\sum_{j=1}^{n}\mathbb{E}|X_j|^p\right)^{1/p}\right)$$

- Johnson, Schechtman and Zinn (1985): best possible $C(p) \asymp \frac{p}{\log(p)}$.

Moment inequalities

- Alternatively, one may ask for $C_1(p)$ and $C_2(p)$ such that

$$\mathbb{E}^{1/p}\Big|\sum_{j=1}^n X_j\Big|^p \le C_1(p)\left(\sum_{j=1}^n \mathbb{E}X_j^2\right)^{1/2} + C_2(p)\left(\mathbb{E}\max_j |X_j|^p\right)^{1/p} \qquad (*)$$

Pinelis, Utev: $C_1(p) \asymp \sqrt{p}$, $C_2(p) \asymp p$ are not improvable in general.

Moment inequalities

- Alternatively, one may ask for $C_1(p)$ and $C_2(p)$ such that

$$\mathbb{E}^{1/p}\Big|\sum_{j=1}^{n} X_j\Big|^p \leq C_1(p)\left(\sum_{j=1}^{n}\mathbb{E}X_j^2\right)^{1/2} + C_2(p)\left(\mathbb{E}\max_j |X_j|^p\right)^{1/p} \qquad (*)$$

  Pinelis, Utev: $C_1(p) \asymp \sqrt{p}$, $C_2(p) \asymp p$ are not improvable in general.

- Nagaev, Pinelis (1977): $\boxed{\text{just integrate (a variant of) Fuk-Nagaev inequality!}}$

Moment inequalities

- Alternatively, one may ask for $C_1(p)$ and $C_2(p)$ such that

$$\mathbb{E}^{1/p}\Big|\sum_{j=1}^{n} X_j\Big|^p \le C_1(p)\left(\sum_{j=1}^{n}\mathbb{E}X_j^2\right)^{1/2} + C_2(p)\left(\mathbb{E}\max_j |X_j|^p\right)^{1/p} \qquad (*)$$

  Pinelis, Utev: $C_1(p) \asymp \sqrt{p}$, $C_2(p) \asymp p$ are not improvable in general.

- Nagaev, Pinelis (1977): $\boxed{\text{just integrate (a variant of) Fuk-Nagaev inequality!}}$
- Up to absolute constants, $(*)$ implies Bernstein's inequality: if $|X_j| \le U$ a.s. for all j, then

$$\mathbb{P}\left(\Big|\sum_{j=1}^{n} X_j\Big| \ge t\right) \le 2\exp\left(-\frac{t^2/2}{\sum_j \mathbb{E}X_j^2 + Ut/3}\right)$$

Moment inequalities

- Alternatively, one may ask for $C_1(p)$ and $C_2(p)$ such that

$$\mathbb{E}^{1/p}\left|\sum_{j=1}^{n} X_j\right|^p \le C_1(p)\left(\sum_{j=1}^{n}\mathbb{E}X_j^2\right)^{1/2} + C_2(p)\left(\mathbb{E}\max_j |X_j|^p\right)^{1/p} \qquad (*)$$

  Pinelis, Utev: $C_1(p) \asymp \sqrt{p}$, $C_2(p) \asymp p$ are not improvable in general.

- Nagaev, Pinelis (1977): $\boxed{\text{just integrate (a variant of) Fuk-Nagaev inequality!}}$
- Up to absolute constants, $(*)$ implies Bernstein's inequality: if $|X_j| \le U$ a.s. for all j, then

$$\mathbb{P}\left(\left|\sum_{j=1}^{n} X_j\right| \ge t\right) \le 2\exp\left(-\frac{t^2/2}{\sum_j \mathbb{E}X_j^2 + Ut/3}\right)$$

  Indeed,

$$\mathbb{P}\left(\left|\sum_{j=1}^{n} X_j\right| \ge C\cdot e\left(\sqrt{p}\left(\sum_{j=1}^{n}\mathbb{E}X_j^2\right)^{1/2} + pU\right)\right)$$

Moment inequalities

- Alternatively, one may ask for $C_1(p)$ and $C_2(p)$ such that

$$\mathbb{E}^{1/p}\Big|\sum_{j=1}^n X_j\Big|^p \le C_1(p)\left(\sum_{j=1}^n \mathbb{E}X_j^2\right)^{1/2} + C_2(p)\left(\mathbb{E}\max_j |X_j|^p\right)^{1/p} \qquad (*)$$

  Pinelis, Utev: $C_1(p) \asymp \sqrt{p}$, $C_2(p) \asymp p$ are not improvable in general.
- Nagaev, Pinelis (1977): $\boxed{\text{just integrate (a variant of) Fuk-Nagaev inequality!}}$
- Up to absolute constants, $(*)$ implies Bernstein's inequality: if $|X_j| \le U$ a.s. for all j, then

$$\mathbb{P}\left(\Big|\sum_{j=1}^n X_j\Big| \ge t\right) \le 2\exp\left(-\frac{t^2/2}{\sum_j \mathbb{E}X_j^2 + Ut/3}\right)$$

  Indeed,

$$\mathbb{P}\left(\Big|\sum_{j=1}^n X_j\Big| \ge C\cdot e\left(\sqrt{p}\left(\sum_{j=1}^n \mathbb{E}X_j^2\right)^{1/2} + pU\right)\right) \le e^{-p}\underbrace{\frac{\mathbb{E}\Big|\sum_{j=1}^n X_j\Big|^p}{C^p\left(\sqrt{p}\left(\sum_{j=1}^n \mathbb{E}X_j^2\right)^{1/2} + pU\right)^p}}_{\le 1}$$

Moment inequalities

- Alternatively, one may ask for $C_1(p)$ and $C_2(p)$ such that

$$\mathbb{E}^{1/p}\Big|\sum_{j=1}^n X_j\Big|^p \le C_1(p)\left(\sum_{j=1}^n \mathbb{E}X_j^2\right)^{1/2} + C_2(p)\left(\mathbb{E}\max_j |X_j|^p\right)^{1/p} \qquad (*)$$

  Pinelis, Utev: $C_1(p) \asymp \sqrt{p},\ C_2(p) \asymp p$ are not improvable in general.
- Nagaev, Pinelis (1977): $\boxed{\text{just integrate (a variant of) Fuk-Nagaev inequality!}}$
- Up to absolute constants, $(*)$ implies Bernstein's inequality: if $|X_j| \le U$ a.s. for all j, then

$$\mathbb{P}\left(\Big|\sum_{j=1}^n X_j\Big| \ge t\right) \le 2\exp\left(-\frac{t^2/2}{\sum_j \mathbb{E}X_j^2 + Ut/3}\right)$$

- In this talk, we will show that it is possible to have $C_1(p) \asymp \sqrt{p},\ C_2(p) \asymp \frac{p}{\log(p)}$ if

$$\mathbb{E}\max_j |X_j| \lesssim (\log(p))^{-1}\left(\mathbb{E}\max_j |X_j|^p\right)^{1/p}.$$

- Let $X_1, \ldots, X_n \in \mathbb{C}^{d \times d}$ be independent, centered, self-adjoint random matrices.

- Let $X_1, \ldots, X_n \in \mathbb{C}^{d\times d}$ be independent, centered, self-adjoint random matrices.
- How large is the spectral norm $\left\|\sum_{j=1}^n X_j\right\|$?

- Let $X_1, \dots, X_n \in \mathbb{C}^{d\times d}$ be independent, centered, self-adjoint random matrices.
- How large is the spectral norm $\left\|\sum_{j=1}^n X_j\right\|$?
- Example: $Y_1, \dots, Y_n \in \mathbb{R}^d$, $\mathbb{E} Y_j = 0$, $\mathbb{E} Y_j Y_j^T = \Sigma$, and $X_j = \frac{1}{n}\left(Y_j Y_j^T - \Sigma\right)$:

$$\Bigg\| \underbrace{\frac{1}{n}\sum_{j=1}^n Y_j Y_j^T}_{:=\widehat{\Sigma}_n} - \Sigma \Bigg\| \leq ?$$

- Let $X_1, \ldots, X_n \in \mathbb{C}^{d\times d}$ be independent, centered, self-adjoint random matrices.
- How large is the spectral norm $\left\|\sum_{j=1}^n X_j\right\|$?
- Example: $Y_1, \ldots, Y_n \in \mathbb{R}^d$, $\mathbb{E}Y_j = 0$, $\mathbb{E}Y_jY_j^T = \Sigma$, and $X_j = \frac{1}{n}\left(Y_jY_j^T - \Sigma\right)$:

$$\Bigg\| \underbrace{\frac{1}{n}\sum_{j=1}^n Y_jY_j^T}_{:=\widehat{\Sigma}_n} - \Sigma \Bigg\| \leq ?$$

- "Matrix Bernstein's" inequality (Ahlswede and Winter; Oliveira; Tropp):
  If $\|X_j\| \leq U$ a.s., then

$$\mathbb{P}\left(\left\|\sum_{j=1}^n X_j\right\| \geq t\right) \leq 2d \exp\left(-\frac{t^2/2}{B_n^2 + Ut/3}\right)$$

  where $B_n^2 \geq \left\|\sum_{j=1}^n \mathbb{E}X_j^2\right\|$ is the so-called "matrix variance."

- "Matrix Bernstein's" inequality (Ahlswede and Winter; Oliveira; Tropp): If $\|X_j\| \le U$ a.s., then

$$\mathbb{P}\left(\left\|\sum_{j=1}^{n} X_j\right\| \ge t\right) \le 2d \exp\left(-\frac{t^2/2}{B_n^2 + Ut/3}\right)$$

  where $B_n^2 \ge \left\|\sum_{j=1}^{n} \mathbb{E}X_j^2\right\|$ is the so-called "matrix variance."

- "Matrix Rosenthal's" inequality: Chen, Gittens, Tropp (2011), Dirksen's Ph.D. thesis (2011), Junge and Zeng (2011). Let $p \ge 2$, then

$$\mathbb{E}^{1/p}\left\|\sum_{j=1}^{n} X_j\right\|^p \lesssim \sqrt{q}B_n + q\mathbb{E}^{1/p} \max_j \|X_j\|^p$$

  where $q = \max(p, \log(ed))$.

- "Matrix Bernstein's" inequality (Ahlswede and Winter; Oliveira; Tropp):
  If $\|X_j\| \le U$ a.s., then

$$\mathbb{P}\left(\left\|\sum_{j=1}^{n} X_j\right\| \ge t\right) \le 2d \exp\left(-\frac{t^2/2}{B_n^2 + Ut/3}\right)$$

  where $B_n^2 \ge \left\|\sum_{j=1}^{n} \mathbb{E}X_j^2\right\|$ is the so-called "matrix variance."
- "Matrix Rosenthal's" inequality: Chen, Gittens, Tropp (2011), Dirksen's Ph.D. thesis (2011), Junge and Zeng (2011). Let $p \ge 2$, then

$$\mathbb{E}^{1/p}\left\|\sum_{j=1}^{n} X_j\right\|^p \lesssim \sqrt{q}B_n + q\mathbb{E}^{1/p}\max_j \|X_j\|^p$$

  where $q = \max(p, \log(ed))$.
- These results are useful tools in statistical applications: matrix completion, community detection, etc.

- "Matrix Bernstein's" inequality (Ahlswede and Winter; Oliveira; Tropp): If $\|X_j\| \le U$ a.s., then

$$\mathbb{P}\left(\left\|\sum_{j=1}^{n} X_j\right\| \ge t\right) \le 2d \exp\left(-\frac{t^2/2}{B_n^2 + Ut/3}\right)$$

where $B_n^2 \ge \left\|\sum_{j=1}^{n} \mathbb{E}X_j^2\right\|$ is the so-called "matrix variance."

- "Matrix Rosenthal's" inequality: Chen, Gittens, Tropp (2011), Dirksen's Ph.D. thesis (2011), Junge and Zeng (2011). Let $p \ge 2$, then

$$\mathbb{E}^{1/p}\left\|\sum_{j=1}^{n} X_j\right\|^p \lesssim \sqrt{q}B_n + q\mathbb{E}^{1/p} \max_j \|X_j\|^p$$

where $q = \max(p, \log(ed))$.

- These results are useful tools in statistical applications: matrix completion, community detection, etc.

- <u>Covariance estimation:</u> if $\mathbb{E}^{1/p} \max_j \|Y_j\|^{2p} \asymp d\|\Sigma\|$, then $\left\|\widehat{\Sigma}_n - \Sigma\right\| \le \varepsilon\|\Sigma\|$ as long as

$$n \gtrsim \frac{1}{\varepsilon^2} d \log(d).$$

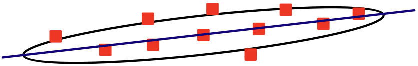
Boundedness and the effective rank

- Covariance estimation: what if Σ can be well approximated by a matrix with small rank? And what if $\|Y_j\|$ is unbounded (but, say, $\mathbb{E}e^{\lambda\|Y_j\|^2} < \infty$ for some $\lambda > 0$)?

Boundedness and the effective rank

- Covariance estimation: what if Σ can be well approximated by a matrix with small rank? And what if $\|Y_j\|$ is unbounded (but, say, $\mathbb{E}e^{\lambda\|Y_j\|^2} < \infty$ for some $\lambda > 0$)?
- Effective rank:

$$r(A) = \frac{\mathrm{tr}(A)}{\|A\|}$$



Boundedness and the effective rank

- Covariance estimation: what if Σ can be well approximated by a matrix with small rank? And what if $\|Y_j\|$ is unbounded (but, say, $\mathbb{E}e^{\lambda\|Y_j\|^2} < \infty$ for some $\lambda > 0$)?
- Effective rank:

$$r(A) = \frac{\mathrm{tr}(A)}{\|A\|}$$

- Results in this direction were obtained by M. (2011), Koltchinskii (2011), Klochkov and Zhivotovskiy (2018). Assume that $\sum_{j=1}^n \mathbb{E}X_j^2 \preceq V_n$, $B_n^2 = \|V_n\|$, $M := \max_j \|X_j\|$. Then

$$\mathbb{P}\left(\left\|\sum_{j=1}^n X_j\right\| \geq t\right) \leq C_1 \cdot r(V_n) \exp\left(-C_2 \frac{t^2}{B_n^2 + \|M\|_{\psi_1} t}\right)$$

Main results: Fuk-Nagaev-type inequality

Theorem (J+M+S+W)

Let $X_1, \ldots, X_n$ be centered self-adjoint random matrices such that $\mathbb{E}\|X_1\|^p < \infty$ and

$$\sum_{j=1}^{n} \mathbb{E}X_j^2 \preceq V_n, \quad B_n^2 = \|V_n\|, \quad M := \max_j \|X_j\|$$

Then

$$\mathbb{P}\left(\left\|\sum_{j=1}^{n} X_j\right\| \geq t\right) \lesssim r(V_n) \exp\left(-C_2 \frac{t^2}{B_n^2 + t\,\mathbb{E}M}\right) + \mathbb{P}(M \geq t/80)$$

$$+\left(\frac{p}{\log(ep)}\right)^{2p} \left(\frac{\mathbb{E}M^p}{t^p}\right)^2$$

Main results: Rosenthal-type inequality

Integrating the Fuk-Nagaev inequality, we deduce the following

Corollary

Let

$$\sum_{j=1}^{n} \mathbb{E} X_j^2 \preceq V_n, \quad B_n^2 = \|V_n\|, \quad M := \max_j \|X_j\|$$

For all $p \geq 1$*,*

$$\mathbb{E}^{1/p} \left\| \sum_{j=1}^{n} X_j \right\|^p \lesssim \sqrt{q} B_n + q\,\mathbb{E} M + \frac{p}{\log(ep)} \left(\mathbb{E} M^p\right)^{1/p} \tag{$*$}$$

where $q = \max(p, \log(r(V_n)))$.

Main results: Rosenthal-type inequality

Integrating the Fuk-Nagaev inequality, we deduce the following

Corollary

Let

$$\sum_{j=1}^{n} \mathbb{E}X_j^2 \preceq V_n, \quad B_n^2 = \|V_n\|, \quad M := \max_j \|X_j\|$$

For all $p \geq 1$*,*

$$\mathbb{E}^{1/p} \left\| \sum_{j=1}^{n} X_j \right\|^p \lesssim \sqrt{q} B_n + q\,\mathbb{E}M + \frac{p}{\log(ep)} \left(\mathbb{E}M^p\right)^{1/p} \qquad (*)$$

where $q = \max(p, \log(r(V_n)))$*.*

- Compare to

$$\mathbb{E}^{1/p} \left\| \sum_{j=1}^{n} X_j \right\|^p \lesssim \sqrt{q} B_n + q \left(\mathbb{E}M^p\right)^{1/p},$$

  $q = \max(p, \log(ed))$.
- If $\mathbb{E}M \ll \mathbb{E}^{1/p} M^p$ then $(*)$ improves the scalar version of Rosenthal's inequality.

Applications to covariance estimation

- $\widehat{\Sigma}_n = \frac{1}{n}\sum_{j=1}^{n} Y_j Y_j^T$, and $\left\|\widehat{\Sigma}_n - \Sigma\right\| \le \varepsilon\|\Sigma\|$ as long as

$$n \gtrsim \frac{1}{\varepsilon^2} d \log(d)$$

Questions: (a) can $\log(d)$ be removed? (b) is $n \gtrsim \frac{1}{\varepsilon^2} r(\Sigma)$ sufficient?

Applications to covariance estimation

- $\widehat{\Sigma}_n = \frac{1}{n}\sum_{j=1}^n Y_j Y_j^T$, and $\left\|\widehat{\Sigma}_n - \Sigma\right\| \leq \varepsilon\|\Sigma\|$ as long as

$$n \gtrsim \frac{1}{\varepsilon^2} d \log(d)$$

  Questions: (a) can $\log(d)$ be removed? (b) is $n \gtrsim \frac{1}{\varepsilon^2} r(\Sigma)$ sufficient?

- Yes! - for Gaussian covariance operators, Koltchinskii and Lounici obtained very general, optimal bounds.

Applications to covariance estimation

- $\widehat{\Sigma}_n = \frac{1}{n}\sum_{j=1}^{n} Y_j Y_j^T$, and $\left\|\widehat{\Sigma}_n - \Sigma\right\| \le \varepsilon\|\Sigma\|$ as long as

$$n \gtrsim \frac{1}{\varepsilon^2} d\log(d)$$

  Questions: (a) can $\log(d)$ be removed? (b) is $n \gtrsim \frac{1}{\varepsilon^2} r(\Sigma)$ sufficient?

- Yes! - for Gaussian covariance operators, Koltchinskii and Lounici obtained very general, optimal bounds.
- For the log-concave and heavy-tailed cases, results by Bourgain; Rudelson; Vershynin, Srivastava; Adamczak, Litvak, Pajor, Tomczak-Jaegermann, Guédon; Mendelson, Paouris.

Applications to covariance estimation

- $\widehat{\Sigma}_n = \frac{1}{n}\sum_{j=1}^n Y_j Y_j^T$, and $\left\|\widehat{\Sigma}_n - \Sigma\right\| \le \varepsilon\|\Sigma\|$ as long as

$$n \gtrsim \frac{1}{\varepsilon^2} d \log(d)$$

  Questions: (a) can $\log(d)$ be removed? (b) is $n \gtrsim \frac{1}{\varepsilon^2} r(\Sigma)$ sufficient?

- Yes! - for Gaussian covariance operators, Koltchinskii and Lounici obtained very general, optimal bounds.
- For the log-concave and heavy-tailed cases, results by Bourgain; Rudelson; Vershynin, Srivastava; Adamczak, Litvak, Pajor, Tomczak-Jaegermann, Guédon; Mendelson, Paouris.
- Tikhomirov (2016): let $\Sigma = I_d$ and assume that $\sup_v \mathbb{E}\,|\langle Y_1, v\rangle|^p < \infty$ where $p > 4$. Then with probability at least $1 - 1/n$,

$$\left\|\widehat{\Sigma}_n - \Sigma\right\| \lesssim_p \frac{\max_j \|Y_j\|^2}{n} + \sqrt{\frac{d}{n}}$$

Applications to covariance estimation

- Tikhomirov (2016): let $\Sigma = I_d$ and assume that $\sup_v \mathbb{E}\, |\langle Y_1, v\rangle|^p < \infty$ where $p > 4$. Then with probability at least $1 - 1/n$,

$$\left\| \widehat{\Sigma}_n - \Sigma \right\| \lesssim_p \frac{\max_j \|Y_j\|^2}{n} + \sqrt{\frac{d}{n}}$$

- Building on results of Abdalla and Zhivotovskiy (2023), one can show the following: assume that $r(\Sigma) < cn$ and that for some $p > 4$,

$$\mathbb{E}^{1/p} \, |\langle Y, v\rangle|^p \le \kappa \mathbb{E}^{1/2} \langle Y, v\rangle^2$$

Then with probability at least $1 - 1/n$,

$$\left\| \widehat{\Sigma}_n - \Sigma \right\| \lesssim_{p,\kappa} \frac{\max_j \|Y_j\|^2}{n} + \|\Sigma\| \sqrt{\frac{r(\Sigma)}{n}}$$

Applications to covariance estimation

- Building on results of Abdalla and Zhivotovskiy (2023), one can show the following: assume that $r(\Sigma) < cn$ and that for some $p > 4$,

$$\mathbb{E}^{1/p} \left|\langle Y, v\rangle\right|^p \leq \kappa \mathbb{E}^{1/2} \langle Y, v\rangle^2$$

Then with probability at least $1 - 1/n$,

$$\left\|\widehat{\Sigma}_n - \Sigma\right\| \lesssim_{p,\kappa} \frac{\max_j \|Y_j\|^2}{n} + \|\Sigma\| \sqrt{\frac{r(\Sigma)}{n}}$$

- What about bounds in L_p? Using Rosenthal's inequality, we prove the following bound:

Corollary

$$\mathbb{E}^{1/2} \left\|\widehat{\Sigma}_n - \Sigma\right\|^2 \lesssim_{p,\kappa} \frac{\left(\mathbb{E} \max_j \|Y_j\|^p\right)^{2/p}}{n} + \|\Sigma\| \sqrt{\frac{r(\Sigma)}{n}}$$

Applications to covariance estimation

- Assume that $r(\Sigma) < cn$ and that for some $p > 4$,

$$\mathbb{E}^{1/p} \left| \langle Y, v \rangle \right|^p \leq \kappa \mathbb{E}^{1/2} \langle Y, v \rangle^2$$

Applications to covariance estimation

- Assume that $r(\Sigma) < cn$ and that for some $p > 4$,

$$\mathbb{E}^{1/p} |\langle Y, v\rangle|^p \leq \kappa \mathbb{E}^{1/2} \langle Y, v\rangle^2$$

- Let $u_j, \hat{u}_j$ be the eigenvectors of $\Sigma, \widehat{\Sigma}_n$,

$$g_1 = \lambda_1 - \lambda_2 \text{ and } g_j = \min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})$$

Applications to covariance estimation

- Assume that $r(\Sigma) < cn$ and that for some $p > 4$,

$$\mathbb{E}^{1/p} |\langle Y, v\rangle|^p \le \kappa \mathbb{E}^{1/2} \langle Y, v\rangle^2$$

- Let $u_j, \hat{u}_j$ be the eigenvectors of $\Sigma, \widehat{\Sigma}_n$,

$$g_1 = \lambda_1 - \lambda_2 \text{ and } g_j = \min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})$$

- "Relative rank"

$$r_j(\Sigma) = \sum_{i \neq j} \frac{\lambda_i}{|\lambda_i - \lambda_j|} + \frac{\lambda_j}{g_j}$$

Applications to covariance estimation

- Assume that $r(\Sigma) < cn$ and that for some $p > 4$,

$$\mathbb{E}^{1/p} |\langle Y, v\rangle|^p \le \kappa \mathbb{E}^{1/2} \langle Y, v\rangle^2$$

- Let $u_j, \hat{u}_j$ be the eigenvectors of $\Sigma, \widehat{\Sigma}_n$,

$$g_1 = \lambda_1 - \lambda_2 \text{ and } g_j = \min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})$$

- "Relative rank"

$$r_j(\Sigma) = \sum_{i \neq j} \frac{\lambda_i}{|\lambda_i - \lambda_j|} + \frac{\lambda_j}{g_j}$$

Corollary

$$\mathbb{E}^{1/2} \left\| \hat{u}_j - u_j \right\|_2^2 \lesssim_{p,\kappa} \sqrt{\frac{\lambda_j}{g_j}} \sqrt{\frac{r_j(\Sigma)}{n} + \frac{r_j(\Sigma)}{n^{1-2/p}}}$$

Idea of the proof of Fuk-Nagaev inequality

- $X_1, \ldots, X_n$ are centered self-adjoint random matrices such that $\mathbb{E}\|X_1\|^p < \infty$ and $M := \max_j \|X_j\|$.

$$\mathbb{P}\left(\left\|\sum_{j=1}^{n} X_j\right\| \geq t\right) \leq ?$$

Idea of the proof of Fuk-Nagaev inequality

- $X_1, \ldots, X_n$ are centered self-adjoint random matrices such that $\mathbb{E}\|X_1\|^p < \infty$ and $M := \max_j \|X_j\|$.

$$\mathbb{P}\left(\left\| \sum_{j=1}^{n} X_j \right\| \geq t \right) \leq ?$$

- "Double truncation:" set $U \asymp \mathbb{E}M$, $y = ct$ and

$$\widetilde{X}_j = X_j I\{\|X_j\| \leq U\}, \quad \Delta_j^{(1)} = X_j\, I\{U < \|X_j\| \leq y\}, \quad \Delta_j^{(2)} = X_j\, I\{\|X_j\| > y\}$$

Idea of the proof of Fuk-Nagaev inequality

- $X_1, \ldots, X_n$ are centered self-adjoint random matrices such that $\mathbb{E}\|X_1\|^p < \infty$ and $M := \max_j \|X_j\|$.

$$\mathbb{P}\left(\left\|\sum_{j=1}^n X_j\right\| \geq t\right) \leq ?$$

- "Double truncation:" set $U \asymp \mathbb{E}M$, $y = ct$ and

$$\widetilde{X}_j = X_j I\{\|X_j\| \leq U\}, \quad \Delta_j^{(1)} = X_j\, I\{U < \|X_j\| \leq y\}, \quad \Delta_j^{(2)} = X_j\, I\{\|X_j\| > y\}$$

- Some work is required to prove that

$$\mathbb{P}\left(\left\|\sum_{k=1}^n \Delta_j^{(1)}\right\| > t/16\right) \lesssim \left(\frac{p}{\log(p)}\right)^{2p} \left(\frac{\mathbb{E}M^p}{t^p}\right)^2$$

  Main tools are the Hoffmann-Jørgensen + Talagrand's comparison inequalities between sums and maxima.

Thank you for listening!